

RESEARCH CENTRE

**Nancy - Grand Est**

IN PARTNERSHIP WITH:

**CNRS, Université de Lorraine**

2021

ACTIVITY REPORT

Project-Team

BIGS

**Biology, genetics and statistics**

IN COLLABORATION WITH: Institut Elie Cartan de Lorraine (IECL)

**DOMAIN**

**Digital Health, Biology and Earth**

**THEME**

**Computational Biology**

# Contents

<b>Project-Team BIGS</b>	<b>1</b>
<b>1 Team members, visitors, external collaborators</b>	<b>2</b>
<b>2 Overall objectives</b>	<b>3</b>
<b>3 Research program</b>	<b>3</b>
3.1 Introduction	3
3.2 Stochastic modeling	3
3.3 Estimation and control for stochastic processes	3
3.4 Algorithms and estimation for graph data	4
3.5 Regression and machine learning	4
<b>4 Application domains</b>	<b>5</b>
4.1 Tumor growth-oncology	5
4.2 Genomic data and micro-organisms population	5
4.3 Epidemiology and e-health	5
4.4 Dynamics of telomeres	5
<b>5 Social and environmental responsibility</b>	<b>5</b>
<b>6 Highlights of the year</b>	<b>6</b>
<b>7 New software and platforms</b>	<b>6</b>
7.1 New software	6
7.1.1 ARMADA	6
7.1.2 cvmgof	6
7.1.3 Harissa	7
<b>8 New results</b>	<b>7</b>
8.1 Stochastic modeling	7
8.1.1 Reconstruction of epigenetic landscapes from single-cell data	7
8.1.2 Modeling and estimation of circulating tumor DNA (ctDNA) dynamics for detecting resistance to targeted therapies	7
8.1.3 Quasi-stationary distributions	8
8.1.4 Evolutionary models of food webs	8
8.1.5 Adaptive dynamics in biological populations	8
8.2 Optimal control of Markov processes	9
8.3 Regression and machine learning	9
8.3.1 Cramér–von Mises goodness-of-fit tests in regression models	9
8.3.2 Online data analysis	10
8.3.3 Change-point detection thresholds in the sequential context	10
8.4 Statistical learning and application in health	11
8.4.1 Analysis of diffuse low-grade gliomas growth	11
8.4.2 Estimation of reference curves for fetal weight	11
8.4.3 Construction of parsimonious event risk scores by an ensemble method. An illustration for short-term predictions in chronic heart failure patients from the GISSI-HF trial	12
8.4.4 Prediction of silencing experiments on gene networks for chronic lymphocytic leukemia	12
8.4.5 A statistical methodology to select covariates in high-dimensional data under dependence. Application to the classification of genetic profiles in oncology	12
8.4.6 Projects linked with the COVID 19 pandemic	12

<b>9</b>	<b>Bilateral contracts and grants with industry</b>	<b>13</b>
9.1	Bilateral contracts with industry	13
<b>10</b>	<b>Partnerships and cooperations</b>	<b>13</b>
10.1	International initiatives	13
10.1.1	Participation in International Programs	13
10.2	National initiatives	14
10.3	Regional initiatives	14
<b>11</b>	<b>Dissemination</b>	<b>14</b>
11.1	Promoting scientific activities	14
11.1.1	Scientific events: organisation	14
11.1.2	Scientific events: selection	15
11.1.3	Journal	15
11.1.4	Invited talks	15
11.1.5	Leadership within the scientific community	15
11.1.6	Scientific expertise	15
11.1.7	Research administration	16
11.2	Teaching - Supervision - Juries	16
11.2.1	Teaching	16
11.2.2	Supervision	18
11.2.3	Juries	19
11.3	Popularization	19
11.3.1	Education	19
11.3.2	Interventions	19
<b>12</b>	<b>Scientific production</b>	<b>19</b>
12.1	Publications of the year	19
12.2	Other	22
12.3	Cited publications	22

## Project-Team BIGS

*Creation of the Project-Team: 2011 January 01*

### Keywords

#### Computer sciences and digital sciences

- A3.1. – Data
- A3.2. – Knowledge
  - A3.2.3. – Inference
- A3.3. – Data and knowledge analysis
  - A3.3.1. – On-line analytical processing
  - A3.3.2. – Data mining
  - A3.3.3. – Big data analysis
- A3.4.1. – Supervised learning
- A3.4.2. – Unsupervised learning
- A3.4.4. – Optimization and learning
- A3.4.7. – Kernel methods
- A6. – Modeling, simulation and control
  - A6.1. – Methods in mathematical modeling
    - A6.1.2. – Stochastic Modeling
  - A6.2. – Scientific computing, Numerical Analysis & Optimization
    - A6.2.3. – Probabilistic methods
    - A6.2.4. – Statistical methods
  - A6.4. – Automatic control
    - A6.4.2. – Stochastic control

#### Other research topics and application domains

- B1. – Life sciences
  - B1.1. – Biology
    - B1.1.2. – Molecular and cellular biology
      - B1.1.10. – Systems and synthetic biology
      - B1.1.11. – Plant Biology
  - B2.2. – Physiology and diseases
    - B2.2.1. – Cardiovascular and respiratory diseases
    - B2.2.3. – Cancer
  - B2.3. – Epidemiology
  - B2.4. – Therapies

## 1 Team members, visitors, external collaborators

### Research Scientists

- Nicolas Champagnat [Team leader, Inria, Senior Researcher, HDR]
- Coralie Fritsch [Inria, Researcher]
- Ulysse Herbach [Inria, Researcher]
- Bruno Scherrer [Inria, Researcher, HDR]

### Faculty Members

- Thierry Bastogne [Univ de Lorraine, Associate Professor, HDR]
- Sandie Ferrigno [Univ de Lorraine, Associate Professor]
- Anne Gégout Petit [Univ de Lorraine, Professor, HDR]
- Jean-Marie Monnez [Univ de Lorraine, Emeritus, HDR]
- Aurélie Muller-Gueudin [Univ de Lorraine, Associate Professor]
- Sophie Mézières [Univ de Lorraine, Associate Professor]
- Pierre Vallois [Univ de Lorraine, Emeritus, HDR]
- Denis Villemonais [Univ de Lorraine, Associate Professor, HDR]

### Post-Doctoral Fellows

- Leo Darrigade [Inria, from Apr 2021]
- Joseph Lam-Weil [Univ de Lorraine, from Jun 2021]
- William Ocafrain [Inria]

### PhD Students

- Vincent Hass [Univ de Lorraine, Inria until Aug 2021, ATER from Sep 2021]
- Rodolphe Loubaton [Univ de Lorraine, ATER]
- Anouk Rago [Univ de Lorraine, from Oct 2021]
- Nassim Sahki [Univ de Lorraine, Inria until Feb 2021, ATER from Mar 2021 until Aug 2021]
- Nino Vieillard [Google, CIFRE]
- Nicolás Zalduendo Vidal [Inria]

### Technical Staff

- Joseph Lam-Weil [Univ de Lorraine, Engineer, from Apr 2021 until Jun 2021]
- Nicolas Thorr [Inria, Engineer, until Jun 2021]

### Administrative Assistant

- Emmanuelle Deschamps [Inria]

## 2 Overall objectives

BIGS is a joint team of Inria, CNRS and Université Lorraine, via the Institut Élie Cartan, UMR 7502 CNRS-UL laboratory in mathematics, of which Inria is a strong partner. One member of BIGS, T. Bastogne, comes from the Research Center of Automatic Control of Nancy (CRAN), with which BIGS has strong relations in the domain "Health-Biology-Signal". Our research is mainly focused on stochastic modeling and statistics but also aiming at a better understanding of biological systems. BIGS involves applied mathematicians whose research interests mainly concern probability and statistics. More precisely, our attention is directed on (1) stochastic modeling, (2) estimation and control for stochastic processes, (3) algorithms and estimation for graph data and (4) regression and machine learning. The main objective of BIGS is to exploit these skills in applied mathematics to provide a better understanding of issues arising in life sciences, with a special focus on (1) tumor growth, (2) photodynamic therapy, (3) population studies of genomic data and of micro-organisms genomics, (4) epidemiology and e-health.

## 3 Research program

### 3.1 Introduction

We give here the main lines of our research that belongs to the domains of probability and statistics. For clarity, we made the choice to structure them in four items. Although this choice was not arbitrary, the outlines between these items are sometimes fuzzy because each of them deals with modeling and inference and they are all interconnected.

### 3.2 Stochastic modeling

Our aim is to propose relevant stochastic frameworks for the modeling and the understanding of biological systems. The stochastic processes are particularly suitable for this purpose. Among them, Markov chains give a first framework for the modeling of population of cells [82, 59]. Piecewise deterministic processes are non diffusion processes also frequently used in the biological context [49, 58, 51]. Among Markov models, we developed strong expertise about processes derived from Brownian motion and Stochastic Differential Equations [76, 57]. For instance, knowledge about Brownian or random walk excursions [83, 74] helps to analyse genetic sequences and to develop inference about them. However, nature provides us with many examples of systems such that the observed signal has a given Hölder regularity, which does not correspond to the one we might expect from a system driven by ordinary Brownian motion.

This situation is commonly handled by noisy equations driven by Gaussian processes such as fractional Brownian motion or fractional fields. The basic aspects of these differential equations are now well understood, mainly thanks to the so-called rough paths tools [66], but also invoking the Russo-Vallois integration techniques [75]. The specific issue of Volterra equations driven by fractional Brownian motion, which is central for the subdiffusion within proteins problem, is addressed in [50]. Many generalizations (Gaussian or not) of this model have been recently proposed for some Gaussian locally self-similar fields, or for some non-Gaussian models [62], or for anisotropic models [44].

### 3.3 Estimation and control for stochastic processes

We develop inference about stochastic processes that we use for modeling. Control of stochastic processes is also a way to optimise administration (dose, frequency) of therapy.

There are many estimation techniques for diffusion processes or coefficients of fractional or multifractional Brownian motion according to a set of observations [61, 40, 48]. However, the inference problem for diffusions driven by a fractional Brownian motion is still in its infancy. Our team has a good expertise about inference of the jump rate and the kernel of piecewise-deterministic Markov processes (PDMP) [39, 35, 38, 37], but there are many directions to go further into. For instance, previous work made the assumption of a complete observation of jumps and mode, which is unrealistic in practice. We tackle the problem of inference of "hidden PDMP". For example, in pharmacokinetics modeling inference, we want to account for the presence of timing noise and identification from longitudinal data. We have expertise on these subjects [41], and we also used mixed models to estimate tumor growth [42].

We consider the control of stochastic processes within the framework of Markov Decision Processes [73] and their generalization known as multi-player stochastic games, with a particular focus on infinite-horizon problems. In this context, we are interested in the complexity analysis of standard algorithms, as well as the proposition and analysis of numerical approximate schemes for large problems in the spirit of [43]. Regarding complexity, a central topic of research is the analysis of the Policy Iteration algorithm, which has made significant progress in the last years [85, 72, 56, 78], but is still not fully understood. For large problems, we have a long experience of sensitivity analysis of approximate dynamic programming algorithms for Markov Decision Processes [79, 81, 77, 65, 80], and we currently investigate whether/how similar ideas may be adapted to multi-player stochastic games.

### 3.4 Algorithms and estimation for graph data

A graph data structure consists of a set of nodes, together with a set of pairs of these nodes called edges. This type of data is frequently used in biology because they provide a mathematical representation of many concepts such as biological structures and networks of relationships in a population. Some attention has recently been focused in the group on modeling and inference for graph data.

Network inference is the process of making inference about the link between two variables, taking into account the information about other variables. [84] gives a very good introduction and many references about network inference and mining. Many methods are available to infer and test edges in Gaussian graphical models [84, 67, 54, 55]. However, the Gaussian assumption does not hold when dealing with typical “zero-inflated” abundance data, and we want to develop inference in this case.

Among graphs, trees play a special role because they offer a good model for many biological concepts, from RNA to phylogenetic trees through plant structures. Our research deals with several aspects of tree data. In particular, we work on statistical inference for this type of data under a given stochastic model. We also work on lossy compression of trees via directed acyclic graphs. These methods enable us to compute distances between tree data faster than from the original structures and with a high accuracy.

### 3.5 Regression and machine learning

Regression models and machine learning aim at inferring statistical links between a variable of interest and covariates. In biological study, it is always important to develop adapted learning methods both in the context of *standard* data and also for data of high dimension (with sometimes few observations) and very massive or online data.

Many methods are available to estimate conditional quantiles and test dependencies [71, 60]. Among them we have developed nonparametric estimation by local analysis via kernel methods [52, 53] and we want to study properties of this estimator in order to derive a measure of risk like confidence band and test. We study also many other regression models like survival analysis, spatio temporal models with covariates. Among the multiple regression models, we want to develop omnibus tests that examine several assumptions together.

Concerning the analysis of high dimensional data, our view on the topic relies on the *French data analysis school*, specifically on Factorial Analysis tools. In this context, stochastic approximation is an essential tool [64], which allows one to approximate eigenvectors in a stepwise manner [69, 68, 70]. BIGS aims at performing accurate classification or clustering by taking advantage of the possibility of updating the information “online” using stochastic approximation algorithms [45]. We focus on several incremental procedures for regression and data analysis like linear and logistic regressions and PCA (Principal Component Analysis).

We also focus on the biological context of high-throughput bioassays in which several hundreds or thousands of biological signals are measured for a posterior analysis. We have to account for the inter-individual variability within the modeling procedure. We aim at developing a new solution based on an ARX (Auto Regressive model with eXternal inputs) model structure using the EM (Expectation-Maximisation) algorithm for the estimation of the model parameters.

## 4 Application domains

### 4.1 Tumor growth-oncology

On this topic, we want to propose branching processes to model the appearance of mutations in tumors, through new collaborations with clinicians who measure a particular quantity called circulating tumor DNA (ctDNA). The final purpose is to use ctDNA as an early biomarker of the resistance to an immunotherapy treatment: it is the aim of the ITMO project. Another topic is the identification of dynamic networks of gene expression. In the ongoing work on low-grade gliomas, a local database of 400 patients will be soon available to construct models. We plan to extend it through national and international collaborations (Montpellier CHU, Montreal CRHUM). Our aim is to build a decision-aid tool for personalised medicine. In the same context, there is a topic of clustering analysis of a brain cartography obtained by sensorial simulations during awake surgery.

### 4.2 Genomic data and micro-organisms population

Despite of his 'G' in the name of BIGS, Genetics is not central in the applications of the team. However, we want to contribute to a better understanding of the correlations between genes through their expression data and of the genetic bases of drug response and disease. We have contributed to methods detecting proteomics and transcriptomics variables linked with the outcome of a treatment.

### 4.3 Epidemiology and e-health

We have many works to do in our ongoing projects in the context of personalized medicine with CHU Nancy. They deal with biomarkers research, prognostic value of quantitative variables and events, scoring, and adverse events. We also want to develop our expertise in rupture detection in a project with APHP (Assistance Publique Hôpitaux de Paris) for the detection of adverse events, earlier than the clinical signs and symptoms. The clinical relevance of predictive analytics is obvious for high-risk patients such as those with solid organ transplantation or severe chronic respiratory disease for instance. The main challenge is the rupture detection in multivariate and heterogeneous signals (for instance daily measures of electrocardiogram, body temperature, spirometry parameters, sleep duration, etc.). Other collaborations with clinicians concern foetopathology and we want to use our work on conditional distribution function to explain fetal and child growth. We have data from the "Service de foetopathologie et de placentologie" of the "Maternité Régionale Universitaire" (CHU Nancy).

### 4.4 Dynamics of telomeres

Telomeres are disposable buffers at the ends of chromosomes which are truncated during cell division; so that, over time, due to each cell division, the telomere ends become shorter. By this way, they are markers of aging. Through a collaboration with Pr A. Benetos, geriatrician at CHU Nancy, we recently obtained data on the distribution of the length of telomeres from blood cells. With members of Inria team TOSCA, we want to work in three connected directions: (1) refine methodology for the analysis of the available data; (2) propose a dynamical model for the lengths of telomeres and study its mathematical properties (long term behavior, quasi-stationarity, etc.); and (3) use these properties to develop new statistical methods. A slot of postdoc position is already planned in the Lorraine Université d'Excellence, LUE project GEENAGE (managed by CHU Nancy).

## 5 Social and environmental responsibility

We followed Inria's recommendations to get involved in the fight against COVID 19. We tried to collaborate with the LCPME laboratory in the purpose to predict the number of SARS-CoV-2 positive patients from the Grand Nancy metropolitan at the Nancy University Hospital from the concentration of SARS-Cov-2 residues in waste water. We have encountered difficulties with the Obépine network in obtaining raw data instead of pre-processed indicators. We made predictions from the incidence rates available on Santé Publique France. The predictions are available on the [siwam website](#).



We were also involved in the **MODCOV19** project, a platform of coordination of research actions about modeling of SARS-CoV-2 (Covid-19) pandemic. We were in particular responsible for the **bibliographic awareness group** of the coordination committee.

## 6 Highlights of the year

The list of permanent members of the team noticeably increased in 2021, due to the arrival of several researchers from the former Inria team TOSCA. These researchers are experts of stochastic modeling and analysis for bio-medical applications. Their arrival led to a strengthening of the first axis of our research program. We are currently proposing a new Inria team SIMBA which takes into account these arrivals and the recent recruitments in the past few years in our team, and more generally on the topic of mathematical biology in Institut Élie Cartan de Lorraine.

## 7 New software and platforms

The team has been developing three new packages.

### 7.1 New software

#### 7.1.1 ARMADA

**Name:** A Statistical Methodology to Select Covariates in High-Dimensional Data under Dependence

**Keywords:** Biostatistics, Aggregated methods, High Dimensional Data, Personalized medicine, Variable selection

**Functional Description:** Two steps variable selection procedure in a context of high-dimensional dependent data with few observations. A first step is dedicated to eliminate the dependency between variables (clustering of variables, followed by factor analysis inside each cluster). A second step consists in variable selection by aggregation of adapted methods.

**News of the Year:** This package is a new one.

**URL:** <https://cran.r-project.org/web/packages/armada/>

**Publications:** [hal-02173568](#), [hal-02363338](#)

**Contact:** Aurélie Muller

**Participants:** Aurélie Muller, Anne Gegout Petit

#### 7.1.2 cvmgof

**Keywords:** Regression, Test, Estimators

**Scientific Description:** Many goodness-of-fit tests have been developed to assess the different assumptions of a (possibly heteroscedastic) regression model. Most of them are "directional" in that they detect departures from a given assumption of the model. Other tests are "global" (or "omnibus") in that they assess whether a model fits a dataset on all its assumptions. `cvmgof` focuses on the task of choosing the structural part of the regression function because it contains easily interpretable information about the studied relationship. It implements 2 nonparametric "directional" tests and one nonparametric "global" test, all based on generalizations of the Cramer-von Mises statistic.

**Functional Description:** `cvmgof` is an R library devoted to Cramer-von Mises goodness-of-fit tests. It implements three nonparametric statistical methods based on Cramer-von Mises statistics to estimate and test a regression model.

**News of the Year:** New version available on CRAN website since Jan 11 2021

**URL:** <https://cran.r-project.org/web/packages/cvmgof/index.html>

**Publication:** hal-03101612

**Contact:** Romain Azais

**Participants:** Sandie Ferrigno, Marie-José Martinez, Romain Azais

### 7.1.3 Harissa

**Name:** Hartree approximation for inference along with a stochastic simulation algorithm

**Keywords:** Gene regulatory networks, Reverse engineering, Molecular simulation

**Functional Description:** Harissa is a Python package for both inference and simulation of gene regulatory networks, based on stochastic gene expression with transcriptional bursting. It was implemented in the context of a mechanistic approach to gene regulatory network inference from single-cell data.

**URL:** <https://github.com/ulyssseherbach/harissa>

**Publications:** hal-03370296, hal-03370228, hal-01646910

**Contact:** Ulysse Herbach

## 8 New results

### 8.1 Stochastic modeling

**Participants:** Nicolas Champagnat, Coralie Fritsch, Anne Gégout-Petit, Vincent Hass, Ulysse Herbach, William Oçafrain, Pierre Vallois, Denis Villemonais, Nicolás Zaldueño Vidal.

#### 8.1.1 Reconstruction of epigenetic landscapes from single-cell data

The aim is to better understand how living cells make decisions (e.g., differentiation of a stem cell into a particular specialized type), seeing decision-making as an emergent property of an underlying complex molecular network. Indeed, it is now proven that cells react probabilistically to their environment: cell types do not correspond to fixed states, but rather to “potential wells” of a certain energy landscape (representing the energy of the possible states of the cell) that we are trying to reconstruct. A first paper proposing a reconstruction method has been submitted [26] in the framework of an international collaboration (USA, Switzerland, France). Another paper is about to be submitted [28], dealing more specifically with the inference of the underlying networks.

*Joint work with Nan Papili Gao (ETH Zurich), Olivier Gandrillon (ENS Lyon), András Páldi (EPHE, Paris), and Rudiyanto Gunawan (University at Buffalo, New York)*

#### 8.1.2 Modeling and estimation of circulating tumor DNA (ctDNA) dynamics for detecting resistance to targeted therapies

Continuation of the ITMO Cancer project, supervised by Nicolas Champagnat, concerning the modeling of circulating tumor DNA (ctDNA) to detect the appearance of resistance to targeted therapies (personalized medicine). After a phase of investigation of possible scenarios in collaboration with Alexandre Harlé of the Institute of Cancerology of Lorraine (ICL), a final model was selected. Based on a mathematical analysis, the members of the project then designed a statistical inference algorithm (learning the parameters of the model, including the genealogical tree of mutations for each patient) which is intended to

be validated on real data currently being acquired at the Nancy CHRU. The general idea is to exploit a “variational principle” that allows to explore the discrete space of family trees, of very large size, through a “pivot” space of continuous parameters, easy to optimize (and in reasonable numbers). A paper detailing the model and its inference is in preparation. The previous method allows for the reconstruction of intratumoral heterogeneity, i.e. the subclone composition of the tumor. Based on these data, we are currently studying models of stochastic tumor growth with an emphasis on interactions between the clones to assess the effects of different treatment strategies.

### 8.1.3 Quasi-stationary distributions

We are continuing our research on quasi-stationary distributions (QSD), that is, distributions of Markov stochastic processes with absorption, which are stationary conditionally on non-absorption. For models of biological populations, absorption corresponds usually to extinction of a (sub-)population. QSDs are fundamental tools to describe the population state before extinction and to quantify the large-time behavior of the probability of extinction.

This year, we solved a general conjecture on the Fleming-Viot particle systems approximating QSDs: in cases where several QSDs exist, it is expected that the stationary distributions of the Fleming-Viot processes approach a particular QSD, called minimal QSD. We proved that this holds true for general absorbed Markov processes with soft obstacles in [7]. We also obtained in [8] criteria based on Lyapunov functions allowing to check general conditions of [47] which characterize the exponential uniform convergence in total variation of conditional distributions of an absorbed Markov process to a unique quasi-stationary distribution. Among the various applications they give, they prove that these conditions apply to any logistic Feller diffusions in any dimension conditioned to the non-extinction of all its coordinates. This question was left partly open since the first work of Cattiaux and Méléard on this topic [46].

Together with M. Benaïm (Univ. Neuchâtel), we studied in [4] stochastic algorithms to approximate quasi-stationary distributions of diffusion processes absorbed at the boundary of a bounded domain. We considered a reinforced version of the diffusion, which is resampled according to its occupation measure when it reaches the boundary. We showed that its occupation measure converges to the unique quasi-stationary distribution of the diffusion process. We also obtained in [24] general criteria ensuring existence, uniqueness and/or exponential convergence properties for quasi-stationary distributions. The criteria were specifically designed to apply to degenerate processes such as hypoelliptic diffusions. We also provided in [25] a counterexample to the uniqueness of a quasi-stationary distribution for a diffusion process which satisfies the weak Hörmander condition.

Together with R. Schott (IECL, Univ. Lorraine), we studied in [6] models of deadlocks in distributed systems, using the approach we developed in [8] to study quasi-stationary distributions, in order to characterize and compute numerically the asymptotic behaviour of the deadlock time and the behaviour of the system before deadlock, both for discrete and for diffusion models.

### 8.1.4 Evolutionary models of food webs

We studied models of food web adaptive evolution in [10]. We identified the biomass conversion efficiency as a key mechanism underlying food web evolution and discussed the relevance of such models to study the evolution of food webs.

*In collaboration with S. Billiard (Univ. Lille).*

### 8.1.5 Adaptive dynamics in biological populations

We studied evolutionary models of bacteria with horizontal transfer in [5]. Horizontal transfer is a common mechanism of DNA exchange between micro-organisms that is thought to be responsible for fast evolution of antibiotic resistance for bacteria or evolution of virulence for pathogens. We considered a scaling of parameters taking into account the influence of negligible but non-extinct populations, allowing us to study specific phenomena observed in these models (re-emergence of traits,

cyclic evolutionary dynamics and evolutionary suicide). This work is done in collaboration with S. Méléard (École Polytechnique) and V.C. Tran (Univ. Paris Est Marne-la-Vallée).

We also worked on general evolutionary models of adaptive dynamics under an assumption of large population and small mutations. This year, we obtained existence, uniqueness and ergodicity results for a centered version of the Fleming-Viot process of population genetics, which is a key step to recover variants of the canonical equation of adaptive dynamics, which describes the long time evolution of the dominant phenotype in the population, under less stringent biological assumptions than in previous works. We plan to complete this work next year.

## 8.2 Optimal control of Markov processes

**Participants:** Bruno Scherrer, Nino Vieillard.

We consider Offline Reinforcement Learning methods. The problem is to learn a policy from logged transitions of an environment, without any interaction. In the presence of function approximation, and under the assumption of limited coverage of the state-action space of the environment, it is necessary to enforce the policy to visit state-action pairs close to the support of logged transitions.

In [17], we propose an iterative procedure to learn a pseudometric (closely related to bisimulation metrics) from logged transitions, and use it to define this notion of closeness. We show its convergence and extend it to the function approximation setting. We then use this pseudometric to define a new lookup based bonus in an actor-critic algorithm: PLOFF. This bonus encourages the actor to stay close, in terms of the defined pseudometric, to the support of logged transitions.

In [18], noticing that an agent in this setting should avoid selecting actions whose consequences cannot be predicted from the data, we take inspiration from the literature on bonus-based exploration to design a new offline RL agent. The core idea is to subtract a prediction-based exploration bonus from the reward, instead of adding it for exploration. This allows the policy to stay close to the support of the dataset. We connect this approach to a more common regularization of the learned policy towards the data. Instantiated with a bonus based on the prediction error of a variational autoencoder, we show that our agent is competitive with the state of the art on a set of continuous control locomotion and manipulation tasks.

*Joint work with Robert Dadashi, Shideh Rezaeifar, Léonard Hussenot, Olivier Pietquin, Olivier Bachem and Matthieu Geist.*

## 8.3 Regression and machine learning

**Participants:** Thierry Bastogne, Sandie Ferrigno, Anne Gégout-Petit, Aurélie Gueudin, Benoît Lalloué, Jean-Marie Monnez, Nassim Sahki, Sophie Wantz-Mézières.

### 8.3.1 Cramér-von Mises goodness-of-fit tests in regression models

Many goodness-of-fit tests have been developed to assess the different assumptions of a (possibly heteroscedastic) regression model. Most of them are 'directional' in that they detect departures from a given assumption of the model. Other tests are 'global' (or 'omnibus') in that they assess whether a model fits a dataset on all its assumptions. We focus on the task of choosing the structural part of the regression function because it contains easily interpretable information about the studied relationship. We consider 2 nonparametric 'directional' tests and one nonparametric 'global' test, all based on generalizations of the Cramér-von Mises statistic.

To perform these goodness-of-fit tests, we develop the R package `cvmgof` [36], an easy-to-use tool for practitioners, available from the Comprehensive R Archive Network (CRAN). The use of the library

is illustrated through a tutorial on real data and simulation studies are carried out in order to show how the package can be exploited to compare the 3 implemented tests. The practitioner can also easily compare the test procedures with different kernel functions, bootstrap distributions, numbers of bootstrap replicates, or bandwidths. The package was updated at the start of 2021, this is its third version. A first article [1] has been published on this work in October 2021.

We are now working on nonparametric tests associated with the functional form of the variance of the regression model. For this, we continue to work on the global test of Ducharme and Ferrigno in order to compare it in terms of performance with directional tests associated with the variance of the model. Many simulations are in progress. This will also make it possible to propose a more general package-type tool making it possible to validate the regression models used in practice.

To complete this work, it would be interesting to assess the other assumptions of a regression model such as the additivity of the random error term. The implementation of these directional tests would enrich the *cvmgof* package and offer a complete easy-to-use tool for validating regression models. Moreover, the assessment of the overall validity of the model when using several directional tests could be compared with that done when using only a global test. In particular, the well-known problem of multiple testing could be discussed by comparing the results obtained from multiple test procedures with those obtained when using a global test strategy. Another perspective of this work would be to develop a similar tool for other statistical models widely used in practice such as generalized linear models.

*Join work with Romain Azaïs (INRIA, ENS Lyon) and Marie-José Martinez (LJK, Université Grenoble Alpes).*

### 8.3.2 Online data analysis

*Widening the scope of an eigenvector stochastic approximation process and application to streaming PCA and related methods.* This article in collaboration with A. Skiredj was presented in the 2020 Activity Report (Section 8.3.5) and is now published in *Journal of Multivariate Analysis* [15].

*Streaming constrained binary logistic regression with online standardized data.* This article in collaboration with E. Albuissou was presented in the 2020 Activity Report (Section 8.3.5) and is now accepted in *Journal of Applied Statistics* [13].

*Construction and update of an online ensemble score involving linear discriminant analysis and logistic regression.* This article in collaboration with E. Albuissou was presented in the 2020 Activity Report (Section 8.3.5) and is being submitted [30], [63].

*Stochastic approximation of eigenvectors and eigenvalues of the  $Q$ -symmetric expectation of a random matrix. Application to streaming PCA.* In this analysis, we have studied the convergence of stochastic approximation processes of the Oja type for estimating eigenvectors of the unknown  $Q$ -symmetric expectation  $B$  of a random matrix, the metric  $Q$  being unknown. We have established a theorem of a.s. convergence of these processes with assumptions on the noisy observations  $B_n$  of  $B$  that are more general than in previous results. The estimation of eigenvectors corresponding to eigenvalues of  $B$  in decreasing order is obtained using at step  $n$  a Gram-Schmidt orthonormalization with respect to a random metric  $Q_{n+1}$  such that  $Q_{n+1}$  converges a.s. to  $Q$  as  $n$  goes to infinity. We have proved the a.s. convergence of specific processes to corresponding eigenvalues. Corollaries of this theorem apply in particular to cases where  $E[B_n|T_n]$  or  $B_n$  converges a.s. to  $B$  which were studied by Monnez and Skiredj [15]. In the case of a process using only the current observations at each step, we have suggested constructing another process using past and current observations. We have applied these results to the online estimation of principal components in streaming PCA of a random vector, taking into account all the observations up to the current step with possibly different weights assigned to past and current observations.

Other applications to methods related to PCA such as generalized canonical correlation analysis are in progress.

### 8.3.3 Change-point detection thresholds in the sequential context

To apply our algorithms of change-point to real data, we turned to some EMG signal data provided by INRS. The study concerns the development of trapezius muscle myalgia in the workplace. We apply change-point detection to characterize different computer activities carried out during an experimental day. Our analysis allowed us to characterize activities according to the frequency and amplitude of

jumps and to distinguish office activities using the mouse from those using the keyboard. This work was presented in a conference paper [19].

## 8.4 Statistical learning and application in health

**Participants:** Nicolas Champagnat, Léo Darrigade, Sandie Ferrigno, Coralie Fritsch, Anne Gégout-Petit, Aurélie Gueudin, Ulysse Herbach, Benoît Lalloué, Rodolphe Loubaton, Jean-Marie Monnez, Anouk Rago, Nicolas Thorr, Pierre Vallois, Sophie Wantz-Mézières.

### 8.4.1 Analysis of diffuse low-grade gliomas growth

In the aim of understanding the growth of low-grade glioma, we investigate multiple fields of information available in clinical practice: patient-related predictors, variables related to tumor tissue and genetics. Monitoring growth through regular MRIs gives us access to many imaging-related variables, including an original one measuring tumor infiltration (thesis defended in 2021: Cyril Brzenczek CRAN, article in preparation). Our last efforts have focused on the statistical analysis of the database composed of these variables. We have obtained a regional fund PACTE to host this database and use it for teaching, dissemination and development of experimentation tools: PIANO platform.

*Join work with J.M. Moureaux, Y. Gaudeau (CRAN), F Rech, L. Taillandier, M. Blonski, T. Obara (CHRU Nancy)*

### 8.4.2 Estimation of reference curves for fetal weight

In Epidemiology, we are working with INSERM to study fetal development in the last two trimesters of pregnancy. Reference or standard curves are required in this kind of biomedical problems. Values that lie outside the limits of these reference curves may indicate the presence of a disorder. Data are from the French EDEN mother-child cohort (INSERM). It's a mother-child cohort study investigating the prenatal and early postnatal determinants of child health and development. 2002 pregnant women were recruited before 24 weeks of amenorrhoea in two maternity clinics from middle-sized French cities (Nancy and Poitiers). From May 2003 to September 2006, 1899 newborns were then included. The main outcomes of interest are fetal (via ultra-sound) and postnatal growth, adiposity development, respiratory health, atopy, behaviour and bone, cognitive and motor development. We are studying fetal weight and height as a function of the gestational age in the third trimester of pregnancy. Some classical empirical and parametric methods such as polynomial regression are first used to construct these curves. For instance, polynomial regression is one of the most common parametric approaches for modeling growth data, especially during the prenatal period. However, these classical methods require strong assumptions. We therefore propose to work with semi-parametric LMS methods, by modifying the response variable (fetal weight) with, among others, Box-Cox transformations. A first article detailing these methodologies applied to the EDEN data should be submitted next year and is the object of the communication [31].

Alternative nonparametric methods as Nadaraya-Watson kernel estimation, local polynomial estimation, B-splines or cubic splines are also developed in this context to construct these curves. The practical implementation of these methods required working on smoothing parameters or choice of knots for the different types of nonparametric estimation. In particular, optimal choice of these parameters has been proposed. Then, a first version of an R package has been developed to propose a tool to construct nonparametric reference curves. It will soon be available on GitHub. In addition, a graphical interface (GUI) intended for practitioners is being developed to allow intuitive visualization of the results given by the package and an article is in progress.

*Join work with Myriam Maumy-Bertrand (IRMA, Université de Strasbourg) and INSERM.*

#### 8.4.3 Construction of parsimonious event risk scores by an ensemble method. An illustration for short-term predictions in chronic heart failure patients from the GISSI-HF trial

This article in collaboration with E. Albuissou and D. Lucci was presented in the 2020 Activity Report (Section 8.4.2) and is now published in *Applied Mathematics* [14].

#### 8.4.4 Prediction of silencing experiments on gene networks for chronic lymphocytic leukemia

We are working with L. Vallat (CHRU Strasbourg) on the inference of dynamical gene networks from RNAseq and proteome data. The goal is to infer a model of gene expression allowing to predict the gene expression in cells where the expression of genes is silenced (e.g. using siRNA), in order to select the silencing experiments which are more likely to reduce the cell proliferation. We expect the selected genes to provide new therapeutic targets for the treatment of chronic lymphocytic leukemia. This year, we addressed the general problem of prediction as defined above, and constructed and proposed an inference method for a new gene network model for which such a prediction is possible. Next year, we expect to identify potential therapeutic targets for which silencing experiments could be conducted.

#### 8.4.5 A statistical methodology to select covariates in high-dimensional data under dependence. Application to the classification of genetic profiles in oncology

We propose a new methodology for selecting and ranking covariates associated with a variable of interest in a context of high-dimensional data under dependence but few observations. The methodology successively intertwines the clustering of covariates, decorrelation of covariates using Factor Latent Analysis, selection using aggregation of adapted methods and finally ranking. A simulation study shows the interest of the decorrelation inside the different clusters of covariates. We first apply our method to transcriptomic data of 37 patients with advanced non-small-cell lung cancer who have received chemotherapy, to select the transcriptomic covariates that explain the survival outcome of the treatment. Secondly, we apply our method to 79 breast tumor samples to define patient profiles for a new metastatic biomarker and associated gene network in order to personalize the treatments. This work is published in [2] and is implemented in R package 'ARMADA'.

*In collaboration with T. Boukhobza and H. Dumond from CRAN, and B. Bastien from biopharmaceutical industry Transgene.*

#### 8.4.6 Projects linked with the COVID 19 pandemic

**Seroprevalence study** Pierre Vallois is the scientific coordinator of the seroprevalence study COVAL Nancy held in Nancy in July 2020 in collaboration with CHRU de Nancy (CIC épidémiologie clinique and Laboratoire de Virologie).

*Background.* The World Health Organisation recommends monitoring the circulation of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2). We aimed to estimate anti-SARS-CoV-2 total immunoglobulin (IgT) antibody seroprevalence and describe symptom profiles and in vitro seroneutralization in Nancy, France, in spring 2020.

*Methods.* Individuals were randomly sampled from electoral lists and invited with household members over 5 years old to be tested for anti-SARS-CoV-2 (IgT, i.e. IgA/IgG/IgM) antibodies by ELISA (Bio-rad). Serum samples were classified according to seroneutralization activity 50 % (NT50) on Vero CCL-81 cells. Age- and sex-adjusted seroprevalence was estimated. Subgroups were compared by chi-square or Fisher exact test and logistic regression.

*Results.* Among 2006 individuals, 43 were SARS-CoV-2-positive; the raw seroprevalence was 2.1 % (95 % confidence interval 1.5 to 2.9), with adjusted metropolitan and national standardized seroprevalence 2.5 % (1.8 to 3.3) and 2.3 % (1.7 to 3.1). Seroprevalence was highest for 20- to 34-year-old participants (4.7 % [2.3 to 8.4]), within than out of socially deprived area (2.5 % vs 1 %,  $P=0.02$ ) and with than without intra-family infection ( $p<10^{-6}$ ). Moreover, 25 % (23 to 27) of participants presented at least one COVID-19 symptom associated with SARS-CoV-2 positivity ( $p<10^{-13}$ ), with anosmia or ageusia highly discriminant (odds ratio 27.8 [13.9 to 54.5]), associated with dyspnea and fever. Among the SARS-CoV-2-positives,

16.3 % (6.8 to 30.7) were asymptomatic. For 31 of these individuals, positive seroneutralization was demonstrated in vitro.

*Conclusions.* In this population of very low anti-SARS-CoV-2 antibody seroprevalence, a beneficial effect of the lockdown can be assumed, with frequent SARS-CoV-2 seroneutralization among IgT-positive patients.

The results were published first in [Medrxiv](#) corresponding to [27] and in a peer-reviewed international journal [11].

**SARS-CoV-2 positive patients in hospital predictions** Participants : A. Gégout-Petit, U. Herbach, N. Thorr.

In collaboration with H. Berry, D. Gemmerlé, T. Lepoutre, D. Maucourt and D. Parsons.

We followed Inria's recommendations to get involved in the fight against COVID 19. We tried to collaborate with the LCPME laboratory in the purpose to predict the number of SARS-CoV-2 positive patients from the Grand Nancy metropolitan at the Nancy University Hospital from the concentration of SARS-Cov-2 residues in waste water. We have encountered difficulties with the Obépine network in obtaining raw data rather than mere indicators. We made predictions from the incidence rates available on Santé Publique France. The predictions are available on the [siwam website](#). Inria hired Nicolas Thorr as engineer during 6 months for this project.

## 9 Bilateral contracts and grants with industry

**Participants:** Bruno Scherrer.

### 9.1 Bilateral contracts with industry

B. Scherrer collaborates with Google Brain on reinforcement learning in the framework of the PhD thesis of Nino Vieillard.

## 10 Partnerships and cooperations

**Participants:** Nicolas Champagnat, Léo Darrigade, Coralie Fritsch, Anne Gégout-Petit, Ulysse Herbach, Joseph Lam-Weil, Jean-Marie Monnez, Aurélie Muller-Gueudin, Pierre Vallois, Denis Villemonnais, Sophie Wantz-Mézières.

### 10.1 International initiatives

#### 10.1.1 Participation in International Programs

**BRN**

**Title:** Biostochastic Research Network

**Partner Institution(s):** • Universidad de Valparaiso (Chile) - CIMFAV – Facultad de Ingeniería - Soledad Torres, Rolando Rebolledo.

- CNRS, Inria & Institut Élie Cartan de Lorraine (France) - N. Champagnat, A. Lejay (coordinator for France), D. Villemonnais, R. Schott.

**Date/Duration:** 2018–2022

**Goal:** Scientific exchange around probabilistic models in population ecology.



## 10.2 National initiatives

- FHU CARTAGE (Fédération Hospitalo Universitaire Cardial and ARterial AGEing). **Leader:** Pr Athanase Benetos. **Participants:** Jean-Marie Monnez, Benoît Lalloué, Anne Gégout-Petit.
- RHU Fight HF (Fighting Heart Failure), located at the University Hospital of Nancy. **Leader:** Pr Patrick Rossignol). **Participants:** Jean-Marie Monnez, Benoît Lalloué.
- ITMO Physics, Mathematics applied to Cancer (2017-2022): “Modeling ctDNA dynamics for detecting targeted therapy resistance”. **Funding organisms:** ITMO Cancer, ITMO Technologies pour la santé de l’alliance nationale pour les sciences de la vie et de la santé (AVIESAN), INCa. **Partners:** Inria and IECL (Institut Élie Cartan de Lorraine), CHRU Strasbourg, CRAN (Centre de Recherche en Automatique de Nancy) and ICL (Institut de Cancérologie de Lorraine). **Leader:** N. Champagnat. **Participants:** L. Darrigade, C. Fritsch, A. Gégout-Petit, U. Herbach, A. Muller-Gueudin, P. Vallois.
- GDR 720 ISIS (funded by CNRS). **Leader:** Laure Blanc-Féraud. **Participant:** Sophie Mézières.

## 10.3 Regional initiatives

- CHRU de Nancy. We have good collaborations with several researchers from CHRU de Nancy. We are involved in LUE Impact Geenage in research axis telomeres.
- Région Grand-Est. In the context of the Telomere project, Anne Gégout-Petit and Denis Villemonais obtained a grant from Grand-Est region to hire Joseph Lam-Weil as a post-doctoral fellow. University of Lorraine and LUE GEENAGE program completed the grant.

## 11 Dissemination

**Participants:** Thierry Bastogne, Nicolas Champagnat, Léo Darrigade, Sandie Ferrigno, Coralie Fritsch, Anne Gégout-Petit, Vincent Hass, Ulysse Herbach, Joseph Lam-Weil, Rodolphe Loubaton, Jean-Marie Monnez, Aurélie Muller-Gueudin, Pierre Vallois, Denis Villemonais, Sophie Wantz-Mézières, Nicolás Zalduendo Vidal.

### 11.1 Promoting scientific activities

#### 11.1.1 Scientific events: organisation

##### Member of the organizing committees

- N. Champagnat, C. Fritsch and U. Herbach organized the workshop **Modélisation de l’hétérogénéité tumorale et thérapies ciblées** (IECL, Univ. Lorraine, 21–22 Oct.).
- N. Champagnat is member of the organizing committee of the conference **A Random Walk in the Land of Stochastic Analysis and Numerical Probability** in honor of Denis Talay (CIRM, Luminy, 3-7 Jan. 2022). Unfortunately, this event is postponed due to sanitary restrictions.
- N. Champagnat was member of the organizing committee of the **Journée Scientifique FCH "Covid"** (Inria Nancy - Grand Est, 28 Sep.).
- U. Herbach co-organized the **IBOMAN 2021 conference** (Institut Curie, Paris, 25-27 Oct.).
- A. Gégout-Petit co-organized the formation session **Journée d’études en statistique : Données manquantes** (SFdS, Fréjus 15-19 Nov.).

### 11.1.2 Scientific events: selection

#### Member of the conference program committees

- N. Champagnat is a member of the scientific committee for the [53èmes Journées de Statistiques](#) (Lyon, 13-17 June 2022).

### 11.1.3 Journal

#### Member of the editorial boards

- N. Champagnat served as co-editor-in-chief with Béatrice Laurent-Bonneau (IMT Toulouse) of *ESAIM: Probability & Statistics* until June. Since then, he served as an associate editor of this journal.
- N. Champagnat serves as an associate editor of *Stochastic Models*.

**Reviewer - reviewing activities.** Here is a selection of the journals for which we regularly write referee reports: *Bernoulli*, *Cell*, *Medicina*, *The Annals of Applied Probability*, *Stochastic Processes and their Applications*, *Journal de Mathématiques Pures et Appliquées*, *ALEA - Latin American Journal of Probability and Mathematical Statistics*, *ESAIM: Probability & Statistics*, *Journal of Theoretical Biology*, *Mathematical Biosciences*, *Journal of Physics A: Mathematical and Theoretical*, *Current Opinion in Systems Biology*, *Bioinformatics*...

### 11.1.4 Invited talks

- N. Champagnat has been invited to give a talk at the online workshop [Biostochastic Networks 2021](#) in November and at the online [Journées Scientifiques du GE2MI sur le thème EDP et modèles biomathématiques](#) in December.
- V. Hass has been invited to give talks at the [École de la Chaire MMB](#) in Aussois in June, at the [Journées de probabilités](#) in Guidel in June and at the [Colloque des Jeunes Probabilistes et Statisticiens \(JPS2021\)](#) in Saint-Pierre d'Oléron in October.
- R. Loubaton has been invited to give a talk at the [Congrès des Jeunes Chercheurs en Mathématiques Appliquées 2021 \(CJC-MA 2021\)](#) in Palaiseau in October.
- N. Zaldueño Vidal has been invited to give a talk at the [Research School of the Chaire Modélisation Mathématique et Biodiversité](#) at Aussois in June and at the Etheridge Group Seminar at Oxford University in June.
- U. Herbach has been invited to give a talk at the [Biohasard 2021](#) online conference in June.

### 11.1.5 Leadership within the scientific community

- A. Gégout-Petit is vice-president of the European Network for Business and Industrial Statistics (ENBIS).

### 11.1.6 Scientific expertise

- C. Fritsch has been a member of the Committee for junior permanent research positions of Inria Nancy - Grand Est.
- A. Gégout-Petit has been a member of several hiring committees: as President for Université Technologique de Compiègne (MCF 26e section); Sorbonne Université (PR 26th section); National jury for 46.1 Professor recruitment; University of Luxembourg, Assistant professor in statistics.

### 11.1.7 Research administration

- N. Champagnat is a member of the coordination committee of **MODCOV19**, a platform of coordination of research actions about modeling of SARS-CoV-2 (Covid-19) pandemic. He heads the bibliographic awareness group.
- N. Champagnat is a member of the Comité de Centre, the COMIPERS and the Commission Information Scientifique et Technique of Inria Nancy - Grand Est and Responsable Scientifique for the library of Mathematics of the IECL. He is also local correspondent of the COERLE (Comité Opérationnel d'Évaluation des Risques Légaux et Éthiques) for the Inria Research Center of Nancy - Grand Est.
- C. Fritsch is a member of the *Commission du Développement Technologique* of Inria Nancy-Grand Est and of the *Commission du personnel* of IECL. She was member of the *Commission Parité-Égalité* of IECL until August. She is the local Radar correspondent for the Inria Research Center of Nancy - Grand Est.
- A. Gégout-Petit is the head of "Institut Élie Cartan de Lorraine" (mathematics laboratory of Université de Lorraine).

## 11.2 Teaching - Supervision - Juries

### 11.2.1 Teaching

BIGS faculty members have teaching obligations at Université de Lorraine and are teaching at least 192 hours each year. They teach probability and statistics at different levels (Licence, Master, Engineering school). Many of them have pedagogical responsibilities.

- D. Villemonais is the head of the Mathematical Engineering Major of ENSMN, Université de Lorraine, France.
- T. Bastogne is in charge of the research master program "Santé Numérique et Imagerie Médicale" with the Faculty of Medicine, Université de Lorraine, France.
- Master: N. Champagnat, Introduction to Quantitative Finance, 12h, M1, second year of ENSMN, Université de Lorraine, France.
- Master: N. Champagnat, Introduction to Quantitative Finance, 9h, M2, third year of ENSMN, Université de Lorraine, France.
- Master: N. Champagnat, Problèmes inverses, 15h, M1, second year of ENSMN, Université de Lorraine, France.
- Master: S. Ferrigno, Experimental designs, 4.5h, M1, fourth year of EEIGM, Université de Lorraine, France.
- Master: S. Ferrigno, Data analyzing and mining, 63h, M1, second year of ENSMN, Université de Lorraine, France.
- Master: S. Ferrigno, Modeling and forecasting, 43h, M1, second year of ENSMN, Université de Lorraine, France.
- Master: S. Ferrigno, Training projects, 18h, M1/M2, second and third year of ENSMN, Université de Lorraine, France.
- Master: A. Muller-Gueudin, Probability and Statistics, 160h, second year of ENSEM and ENSAIA, Université de Lorraine, France.
- Master: A. Muller-Gueudin, Scientific calculation with Matlab, 20h, second year of ENSAIA, Université de Lorraine, France.

- Master: A. Gégout-Petit, Statistics, modeling, data analysis, 80h, master in applied mathematics, Université de Lorraine, France.
- Master: U. Herbach, Data analyzing and mining tutorial, 18h, M1, second year of ENSMN, Université de Lorraine, France.
- Master: U. Herbach, Introduction to probability theory, 18h, M1, second year of ENSEM (apprenticeship cursus), Université de Lorraine, France.
- Master: R. Loubaton, Analyse de données, 16h, M1, second year of ENSMN, Université de Lorraine, France.
- Master: R. Loubaton, Introduction à l'apprentissage automatique, 6h, M1, second year of ENSMN, Université de Lorraine, France.
- Master: S. Wantz-Mézières, Learning and analysis of medical data, 36h, with J.M. Moureaux, Master SNIM, Université de Lorraine, France.
- Master: D. Villemonais, Probability Theory II, 63h, M1, second year of ENSMN, Université de Lorraine, France.
- Master: D. Villemonais, Stochastic processes, 32h, Master 2 MFA, Université de Lorraine, France.
- Master: D. Villemonais, Modeling and forecasting, 14h, M1, second year of ENSMN, Université de Lorraine, France.
- Licence: S. Wantz-Mézières, Applied mathematics for management, financial mathematics, Probability and Statistics, 160h, IUT Nancy-Charlemagne (L1/L2/L3), Université de Lorraine, France.
- Licence: S. Wantz-Mézières, Probability, 100h, first year in TELECOM Nancy (initial and apprenticeship cursus), Université de Lorraine, France.
- Licence: A. Muller-Gueudin, Statistics, 60h, first year of ENSAIA, Université de Lorraine, France.
- Licence: S. Ferrigno, Descriptive and inferential statistics, 60h, L2, second year of EEIGM, Université de Lorraine, France.
- Licence: S. Ferrigno, Statistical modeling, 60h, L2, second year of EEIGM, Université de Lorraine, France.
- Licence: S. Ferrigno, Mathematical and computational tools, 20h, L3, third year of EEIGM, Université de Lorraine, France.
- Licence: S. Ferrigno, Training projects, 40h, L1/L3, first, second and third year of EEIGM, Université de Lorraine, France.
- Licence: C. Fritsch, Probability Theory tutorial, 40h, L3, first year of ENSMN, Université de Lorraine, France.
- Licence: V. Hass, Complément d'analyse, 38h, L1, FST, Université de Lorraine, France.
- Licence: V. Hass, Analyse numérique et optimisation, 46h, L3, first year of ENSMN, Université de Lorraine, France.
- Licence: V. Hass, Probabilités, 40h, L3, first year of ENSMN, Université de Lorraine, France.
- Licence: V. Hass, Mathématiques FIGIM 1A, 35h, L1/L2, first year of ENSMN, Université de Lorraine, France.
- Licence: V. Hass, Mathématiques FIGIM 2A, 21h, L2, first year of ENSMN, Université de Lorraine, France.
- Licence: U. Herbach, Statistics tutorial, 39h, L3, first year of ENSMN, Université de Lorraine, France.

- Licence: R. Loubaton, Inférence statistique, 21h, L3, first year of ENSMN, Université de Lorraine, France.
- Licence: R. Loubaton, Probabilités, 20h, L2, FST, Université de Lorraine, France.
- Licence: R. Loubaton, FST, Méthodes Numériques, 10h, L2, FST, Université de Lorraine, France.
- Licence: R. Loubaton, Latex, 9h, L2, FST, Université de Lorraine, France.
- Licence: R. Loubaton, Remédiation mathématique, 30h, L3, first year of ENSMN, Université de Lorraine, France.
- Licence: R. Loubaton, Analyse numérique et optimisation, 40h, L3, first year of ENSMN, Université de Lorraine, France.
- Licence: D. Villemonais, Probability Theory, 57h, L3, first year of ENSMN, Université de Lorraine, France.
- Licence: N. Zalduendo Vidal, Probability Theory tutorial, 40h, L3, first year of ENSMN, Université de Lorraine, France.
- Licence: N. Zalduendo Vidal, Numerical Analysis tutorial, 20h, L3, first year of ENSMN, Université de Lorraine, France.

### 11.2.2 Supervision

#### PhD

- PhD: Nassim Sahki, “Data-driven methodology for sequential change-point detection for physiological signals”, grant Inria-Cordis. Defence 29 Nov 2021. Advisors: A. Gégout-Petit, S. Wantz-Mézières [23].
- PhD in progress: Vincent Hass, “Individual-based models in adaptive dynamics and long time evolution under assumptions of rare advantageous mutations”, grant Inria-Cordis. Advisor: N. Champagnat.
- PhD in progress: Rodolphe Loubaton, “Caractérisation des cibles thérapeutiques dans un programme génique tumoral”, grant Région Grand-Est. Advisors: N. Champagnat and L. Vallat (CHRU Strasbourg).
- PhD in progress: Anouk Rago, “Inférence de réseaux de gènes dynamiques et prédiction d’expériences d’interventions biologiques dans des cellules cancéreuses”, grant Région Grand-Est, Inria. Advisors: N. Champagnat, A. Gégout-Petit.
- PhD in progress: Nino Vieillard, “Approximate Dynamic Programming and Deep Reinforcement Learning”, CIFRE with Google Brain. Advisors: B. Scherrer, M. Geist (Google Brain).
- PhD in progress: Nicolás Zalduendo Vidal, “Processus de branchement bi-sexués multi-types”, grant Inria-Cordis. Advisors: C. Fritsch, D. Villemonais.

#### Other

- Parcours Recherche: Asmaa Labtaina, “Processus de Markov déterministes par morceaux et leur application à l’expression stochastique des gènes” (full-year research project, M1 ENSMN). Advisor: U. Herbach.
- TER: Mohammed Khatbane and Abdelkabar Bouyghf, “Méthodes variationnelles en apprentissage statistique : l’exemple du modèle Latent Dirichlet Allocation” (research project, M1 Univ. Lorraine). Advisor: U. Herbach.

### 11.2.3 Juries

- PhD: N. Champagnat, reporter, thesis of Maxime Berger, “Le comportement critique de la quasi-espèce”, Université PSL.
- PhD: N. Champagnat, reporter, thesis of Felipe Munoz-Hernandez, “Approximation quantitative en grande population de modèles stochastiques avec interaction ou environnement variable”, Institut Polytechnique de Paris.
- PhD: N. Champagnat, reporter, thesis of Julie Tourniaire, “Spatial dynamics of interfaces in ecology: deterministic and stochastic models”, Institut Polytechnique de Paris.
- PhD: A. Gégout-Petit, president, thesis of A. Conanec Rago, Université de Bordeaux.
- PhD: A. Gégout-Petit, president, thesis of S. Yacheur, Université de Lorraine.
- PhD Prize: A. Gégout-Petit, jury member, 2021 AMIES PhD Prize.
- PhD: B. Scherrer, reporter, thesis of Yannis Flet-Berliac, “Sample-Efficient Deep Reinforcement Learning for Control, Exploration and Safety”, Université de Lille.
- PhD: B. Scherrer, reporter, thesis of Marc Etheve, “Using machine learning to solve repeated optimization problems”, CNAM Paris (CIFRE with EDF).

## 11.3 Popularization

### 11.3.1 Education

- S. Ferrigno: Advisor of a group of students (EEIGM), "Traitement statistique de données" project, various high schools, Nancy, 2021
- S. Ferrigno: Advisor of a group of students (EEIGM), "La main à la Pâte" project, Institut médico-éducatif (IME), Commercy, October-November 2021
- S. Mézières: organisation of a research training week on Neurooncology and Numerics, for medical and engineering students, January 2021

### 11.3.2 Interventions

- C. Fritsch made two interventions in the Lycée Cormontaigne in Metz, as part of the “Chiche!” program, in December 2021.

## 12 Scientific production

### 12.1 Publications of the year

#### International journals

- [1] R. Azaïs, S. Ferrigno and M.-J. Martinez. ‘cvmgof: an R package for Cramér-von Mises goodness-of-fit tests in regression models’. In: *Journal of Statistical Computation and Simulation* (24th Oct. 2021). DOI: [10.1080/00949655.2021.1991346](https://doi.org/10.1080/00949655.2021.1991346). URL: <https://hal.archives-ouvertes.fr/hal-03101612>.
- [2] B. Bastien, T. Boukhobza, H. Dumond, A. Gégout-Petit, A. Muller-Gueudin and C. Thiébaud. ‘A statistical methodology to select covariates in high-dimensional data under dependence. Application to the classification of genetic profiles in oncology’. In: *Journal of Applied Statistics* (2021), p. 23. DOI: [10.1080/02664763.2020.1837083](https://doi.org/10.1080/02664763.2020.1837083). URL: <https://hal.archives-ouvertes.fr/hal-02173568>.
- [3] T. Bastogne. ‘iQbD: a TRL-indexed quality-by-design paradigm for medical device engineering’. In: *Journal of Medical Devices* (2021). URL: <https://hal.archives-ouvertes.fr/hal-03505533>.

- [4] M. Benaïm, N. Champagnat and D. Villemonais. ‘Stochastic approximation of quasi-stationary distributions for diffusion processes in a bounded domain’. In: *Annales de l’Institut Henri Poincaré (B) Probabilités et Statistiques* 57.2 (2021), pp. 726–739. DOI: [10.1214/20-AIHP1093](https://doi.org/10.1214/20-AIHP1093). URL: <https://hal.archives-ouvertes.fr/hal-02101739>.
- [5] N. Champagnat, S. Méléard and V. C. Tran. ‘Stochastic analysis of emergence of evolutionary cyclic behavior in population dynamics with transfer’. In: *Annals of Applied Probability* 31.4 (2021), pp. 1820–1867. DOI: [10.1214/20-AAP1635](https://doi.org/10.1214/20-AAP1635). URL: <https://hal.inria.fr/hal-01974289>.
- [6] N. Champagnat, R. Schott and D. Villemonais. ‘Analysis of distributed systems via quasi-stationary distributions’. In: *Stochastic Analysis and Applications* 36.6 (2021), pp. 981–998. DOI: [10.1080/07362994.2020.1861952](https://doi.org/10.1080/07362994.2020.1861952). URL: <https://hal.inria.fr/hal-01710663>.
- [7] N. Champagnat and D. Villemonais. ‘Convergence of the Fleming-Viot process toward the minimal quasi-stationary distribution’. In: *ALEA : Latin American Journal of Probability and Mathematical Statistics* 18 (2021), pp. 1–15. DOI: [10.30757/ALEA.v18-01](https://doi.org/10.30757/ALEA.v18-01). URL: <https://hal.archives-ouvertes.fr/hal-01895618>.
- [8] N. Champagnat and D. Villemonais. ‘Lyapunov criteria for uniform convergence of conditional distributions of absorbed Markov processes’. In: *Stochastic Processes and their Applications* 135 (May 2021), pp. 51–74. DOI: [10.1016/j.spa.2020.12.005](https://doi.org/10.1016/j.spa.2020.12.005). URL: <https://hal.inria.fr/hal-01503697>.
- [9] A. M. G. Cox, E. L. Horton, A. E. Kyprianou and D. Villemonais. ‘Stochastic Methods for Neutron Transport Equation III: Generational many-to-one and  $k_{\text{eff}}$ ’. In: *SIAM Journal on Applied Mathematics* 81.3 (27th May 2021). DOI: [10.1137/19M1295854](https://doi.org/10.1137/19M1295854). URL: <https://hal.archives-ouvertes.fr/hal-02390745>.
- [10] C. Fritsch, S. Billiard and N. Champagnat. ‘Identifying conversion efficiency as a key mechanism underlying food webs evolution : a step forward, or backward?’ In: *Oikos* 130.6 (2021), pp. 904–930. DOI: [10.1111/oik.07421](https://doi.org/10.1111/oik.07421). URL: <https://hal.archives-ouvertes.fr/hal-02129313>.
- [11] A. Gégout-Petit, H. Jeulin, K. Legrand, N. Jay, A. Bochnakian, P. Vallois, E. Schvoerer and F. Guillemin. ‘Seroprevalence of SARS-CoV-2, Symptom Profiles and Sero-Neutralization in a Suburban Area, France’. In: *Viruses* 13.6 (June 2021), p. 1076. DOI: [10.3390/v13061076](https://doi.org/10.3390/v13061076). URL: <https://hal.archives-ouvertes.fr/hal-03269207>.
- [12] J.-P. Jehl, P. Dan, A. Voignier, N. Tran, T. Bastogne, P. Maureira and F. Cleymand. ‘Transverse isotropic modelling of left-ventricle passive filling: mechanical characterization for epicardial biomaterial manufacturing’. In: *Journal of the mechanical behavior of biomedical materials* 119 (July 2021), p. 104492. DOI: [10.1016/j.jmbbm.2021.104492](https://doi.org/10.1016/j.jmbbm.2021.104492). URL: <https://hal.univ-lorraine.fr/hal-03189522>.
- [13] B. Lalloué, J.-M. Monnez and E. Albuissou. ‘Streaming constrained binary logistic regression with online standardized data’. In: *Journal of Applied Statistics* (2021). DOI: [10.1080/02664763.2020.1870672](https://doi.org/10.1080/02664763.2020.1870672). URL: <https://hal.archives-ouvertes.fr/hal-02156324>.
- [14] B. Lalloué, J.-M. Monnez, D. Lucci and E. Albuissou. ‘Construction of Parsimonious Event Risk Scores by an Ensemble Method. An Illustration for Short-Term Predictions in Chronic Heart Failure Patients from the GISSI-HF Trial’. In: *Applied Mathematics* 12.7 (21st July 2021), pp. 627–653. DOI: [10.4236/am.2021.127045](https://doi.org/10.4236/am.2021.127045). URL: <https://hal.archives-ouvertes.fr/hal-03040390>.
- [15] J.-M. Monnez and A. Skiredj. ‘Widening the scope of an eigenvector stochastic approximation process and application to streaming PCA and related methods’. In: *Journal of Multivariate Analysis* 182 (Mar. 2021), p. 104694. DOI: [10.1016/j.jmva.2020.104694](https://doi.org/10.1016/j.jmva.2020.104694). URL: <https://hal.inria.fr/hal-03038206>.
- [16] M. Waelli, E. Minvielle, M. X. Acero, K. Ba and B. Lalloué. ‘What matters to patients? A mixed method study of the importance and consideration of oncology patient demands’. In: *BMC Health Services Research* 21 (2021), p. 256. DOI: [10.1186/s12913-021-06247-0](https://doi.org/10.1186/s12913-021-06247-0). URL: <https://hal.earth.es.fr/hal-02563438>.

### International peer-reviewed conferences

- [17] R. Dadashi, S. Rezaeifar, N. Vieillard, L. Hussenot, O. Pietquin and M. Geist. ‘Offline Reinforcement Learning with Pseudometric Learning’. In: 38th International Conference on Machine Learning. Vol. 139. virtual, France, 18th June 2021. URL: <https://hal.inria.fr/hal-03468847>.
- [18] S. Rezaeifar, R. Dadashi, N. Vieillard, L. Hussenot, O. Bachem, O. Pietquin and M. Geist. ‘Offline Reinforcement Learning as Anti-Exploration’. In: 36th AAAI Conference on Artificial Intelligence. Vancouver, Canada, 22nd Feb. 2022. URL: <https://hal.archives-ouvertes.fr/hal-03468875>.
- [19] N. Sahki, A. Gégout-Petit and S. Wantz-Mézières. ‘Detection of breaks in EMG signals of upper trapezius muscle activity’. In: JDS 2021 - 52èmes Journées de Statistique de la SFdS. Nice / Virtual, France, 7th June 2021. URL: <https://hal.inria.fr/hal-03328366>.

### Conferences without proceedings

- [20] T. Bastogne, S. Bevers, S. Kooijmans, L. Hassler, S. E. Andaloussi, R. Schiffelers and S. De Koker. ‘easyQBD: A quality by design SaaS platform. Application to the development of lipid nanoparticles for mRNA delivery.’ In: 6th Bioproduction Congress. Lyon, France, 30th Sept. 2021. URL: <https://hal.archives-ouvertes.fr/hal-03505539>.
- [21] S. Bevers, S. Kooijmans, E. van de Velde, M. Evers, S. Seghers, J. Gitz-François, L. Hassler, K. Breckpot, T. Bastogne, R. Schiffelers and S. De Koker. ‘Tuning LNPs to target antigen presenting cells in spleen induces CD8 T-cell responses and tumor regression in mice’. In: 18th CIMT Annual Meeting. Mainz, Germany, 10th May 2021. URL: <https://hal.univ-lorraine.fr/hal-03297832>.
- [22] S. Kooijmans, S. Bevers, E. van de Velde, M. J. W. Evers, S. Seghers, J. J. J. M. Gitz-François, L. Hassler, K. Breckpot, T. Bastogne, R. M. Schiffelers and S. De Koker. ‘Rationally designed mRNA-loaded lipid nanoparticles provoke strong antitumor T cell immunity which critically depends on specific immune cell subsets’. In: Annual Meeting of the Controlled Release Society, CRS 2021. Virtual, United States, 25th July 2021. URL: <https://hal.univ-lorraine.fr/hal-03298753>.

### Doctoral dissertations and habilitation theses

- [23] N. Sahki. ‘Data-driven methodology for sequential change-point detection for physiological signals’. Université de Lorraine; École doctorale IAEM Lorraine - Informatique, Automatique, Électronique - Électrotechnique, Mathématiques de Lorraine, 29th Nov. 2021. URL: <https://hal.inria.fr/te1-03469586>.

### Reports & preprints

- [24] M. Benaïm, N. Champagnat, W. Oçafrain and D. Villemonais. *Degenerate processes killed at the boundary of a domain*. 2021. URL: <https://hal.inria.fr/hal-03278551>.
- [25] M. Benaïm, N. Champagnat, W. Oçafrain and D. Villemonais. *Transcritical bifurcation for the conditional distribution of a diffusion process*. 9th Dec. 2021. URL: <https://hal.archives-ouvertes.fr/hal-03471911>.
- [26] N. P. Gao, O. Gandrillon, A. Páldi, U. Herbach and R. Gunawan. *Universality of cell differentiation trajectories revealed by a reconstruction of transcriptional uncertainty landscapes from single-cell transcriptomic data*. 5th Feb. 2021. DOI: 10.1101/2020.04.23.056069. URL: <https://hal.inria.fr/hal-03132652>.
- [27] A. Gégout-Petit, H. Jeulin, K. Legrand, A. Bochnakian, P. Vallois, E. Schvoerer and F. Guillemin. *Seroprevalence of SARS-CoV-2, symptom profiles and seroneutralization during the first COVID-19 wave in a suburban area, France*. 23rd June 2021. DOI: 10.1101/2021.02.10.21250862. URL: <https://hal.archives-ouvertes.fr/hal-03269203>.
- [28] U. Herbach. *Gene regulatory network inference from single-cell data using a self-consistent proteomic field*. 7th Oct. 2021. URL: <https://hal.archives-ouvertes.fr/hal-03370228>.



- [29] S. Janson, C. Mailler and D. Villemonais. *Fluctuations of balanced urns with infinitely many colours*. Nov. 2021. URL: <https://hal.archives-ouvertes.fr/hal-03508437>.
- [30] B. Lalloué, J.-M. Monnez and E. Albuissou. *Construction and update of an online ensemble score involving linear discriminant analysis and logistic regression*. 8th Feb. 2021. URL: <https://hal.archives-ouvertes.fr/hal-03134248>.

### Other scientific publications

- [31] S. Ferrigno. ‘Semiparametric reference curves for EDEN cohort’. In: *CMStatistics 2021*. Londres, United Kingdom, 18th Dec. 2021. URL: <https://hal.archives-ouvertes.fr/hal-03507796>.

## 12.2 Other

### Educational activities

- [32] J.-M. Monnez. ‘Cours d’analyse des données et apprentissage : L’analyse en composantes principales’. Master. France, 29th Apr. 2021. URL: <https://hal.archives-ouvertes.fr/hal-03212055>.
- [33] J.-M. Monnez. ‘Méthodes de classification non supervisée’. Master. France, 29th Apr. 2021. URL: <https://hal.archives-ouvertes.fr/hal-03184300>.

### Softwares

- [34] [SW] U. Herbach, *Harissa: tools for mechanistic gene network inference from single-cell data*, 7th Oct. 2021. LIC: BSD 3-Clause "New" or "Revised" License. HAL: [hal-03370296](https://hal.archives-ouvertes.fr/hal-03370296), URL: <https://hal.archives-ouvertes.fr/hal-03370296>, VCS: <https://github.com/ulysseherbach/harissa>, SWHID: [swh:1:dir:a12e221925f0884f70a234c9a02a25545521d099;origin=https://hal.archives-ouvertes.fr/hal-03370296;visit=swh:1:snp:7a981f6f1cd0d276b920f232bbc5d612e245f1d8;anchor=swh:1:rev:29a70ab8d8eadf78863a8b8bd55941b6f840158f;path=/](https://hal.archives-ouvertes.fr/hal-03370296)).

## 12.3 Cited publications

- [35] R. Azaïs, F. Dufour and A. Gégout-Petit. ‘Non-Parametric Estimation of the Conditional Distribution of the Interjumping Times for Piecewise-Deterministic Markov Processes’. In: *Scandinavian Journal of Statistics* 41.4 (Dec. 2014), pp. 950–969. DOI: [10.1111/sjos.12076](https://doi.org/10.1111/sjos.12076). URL: <https://hal.archives-ouvertes.fr/hal-01103700>.
- [36] [SW] R. Azaïs, S. Ferrigno and M.-J. Martinez, *cvmgof: Cramer-von Mises goodness-of-fit tests version 1.0.0*, Nov. 2018. LIC: CeCILL. HAL: [hal-02014516](https://hal.archives-ouvertes.fr/hal-02014516), URL: <https://hal.archives-ouvertes.fr/hal-02014516>, SWHID: [swh:1:dir:6b9e75ac30a78bdf2d928b2177e8edf538c16748](https://hal.archives-ouvertes.fr/hal-02014516)).
- [37] R. Azaïs and A. Muller-Gueudin. ‘Optimal choice among a class of nonparametric estimators of the jump rate for piecewise-deterministic Markov processes’. In: *Electronic journal of statistics* (2016). URL: <https://hal.archives-ouvertes.fr/hal-01168651>.
- [38] R. Azaïs. ‘A recursive nonparametric estimator for the transition kernel of a piecewise-deterministic Markov process’. In: *ESAIM: Probability and Statistics* 18 (2014), pp. 726–749.
- [39] R. Azaïs, F. Dufour and A. Gégout-Petit. ‘Nonparametric estimation of the jump rate for non-homogeneous marked renewal processes’. In: *Annales de l’Institut Henri Poincaré, Probabilités et Statistiques*. Vol. 49. 4. Institut Henri Poincaré. 2013, pp. 1204–1231.
- [40] J. M. Bardet, G. Lang, G. Oppenheim, A. Philippe, S. Stoev and M. Taqqu. ‘Semi-parametric estimation of the long-range dependence parameter: a survey’. In: *Theory and applications of long-range dependence*. Birkhauser Boston, 2003, pp. 557–577.

- [41] T. Bastogne, S. Mézières-Wantz, N. Ramdani, P. Vallois and M. Barberi-Heyob. 'Identification of pharmacokinetics models in the presence of timing noise'. In: *Eur. J. Control* 14.2 (2008), pp. 149–157. DOI: [10.3166/ejc.14.149-157](https://doi.org/10.3166/ejc.14.149-157). URL: <http://dx.doi.org/10.3166/ejc.14.149-157>.
- [42] T. Bastogne, A. Samson, P. Vallois, S. Wantz-Mézières, S. Pinel, D. Bechet and M. Barberi-Heyob. 'Phenomenological modeling of tumor diameter growth based on a mixed effects model'. In: *Journal of theoretical biology* 262.3 (2010), pp. 544–552.
- [43] D. Bertsekas and J. Tsitsiklis. *Neurodynamic Programming*. Athena Scientific, 1996.
- [44] H. Biermé, C. Lacaux and H.-P. Scheffler. 'Multi-operator Scaling Random Fields'. Anglais. In: *Stochastic Processes and their Applications* 121.11 (2011). MAP5 2011-01, pp. 2642–2677. DOI: [10.1016/j.spa.2011.07.002](https://doi.org/10.1016/j.spa.2011.07.002). URL: <http://hal.archives-ouvertes.fr/hal-00551707/en/>.
- [45] H. Cardot, P. Cénac and J.-M. Monnez. 'A fast and recursive algorithm for clustering large datasets with k-medians'. In: *Computational Statistics & Data Analysis* 56.6 (2012), pp. 1434–1449.
- [46] P. Cattiaux and S. Méléard. 'Competitive or weak cooperative stochastic Lotka–Volterra systems conditioned on non-extinction'. In: *Journal of mathematical biology* 60.6 (2010), pp. 797–829.
- [47] N. Champagnat and D. Villemonais. 'Exponential convergence to quasi-stationary distribution and Q-process'. In: *Probability Theory and Related Fields* 164.1 (2016). 46 pages, pp. 243–283. DOI: [10.1007/s00440-014-0611-7](https://doi.org/10.1007/s00440-014-0611-7). URL: <https://hal.archives-ouvertes.fr/hal-00973509>.
- [48] J. F. Coeurjolly. 'Simulation and identification of the fractional brownian motion: a bibliographical and comparative study'. In: *Journal of Statistical Software* 5 (2000), pp. 1–53.
- [49] M. H. Davis. 'Piecewise-deterministic Markov processes: A general class of non-diffusion stochastic models'. In: *Journal of the Royal Statistical Society. Series B (Methodological)* (1984), pp. 353–388.
- [50] A. Deya and S. Tindel. 'Rough Volterra equations. I. The algebraic integration setting'. In: *Stoch. Dyn.* 9.3 (2009), pp. 437–477. DOI: [10.1142/S0219493709002737](https://doi.org/10.1142/S0219493709002737). URL: <http://dx.doi.org/10.1142/S0219493709002737>.
- [51] M. Doumic, M. Hoffmann, N. Krell and L. Robert. 'Statistical estimation of a growth-fragmentation model observed on a genealogical tree'. In: *Bernoulli* 21.3 (2015), pp. 1760–1799.
- [52] S. Ferrigno and G. Ducharme. 'Un test d'adéquation global pour la fonction de répartition conditionnelle'. In: *C. R. Math. Acad. Sci. Paris* 341.5 (2005), pp. 313–316. DOI: [10.1016/j.crma.2005.07.003](https://doi.org/10.1016/j.crma.2005.07.003). URL: <http://dx.doi.org/10.1016/j.crma.2005.07.003>.
- [53] S. Ferrigno, M. Maumy-Bertrand and A. Muller-Gueudin. 'Uniform law of the logarithm for the local linear estimator of the conditional distribution function'. In: *C. R. Math. Acad. Sci. Paris* 348.17-18 (2010), pp. 1015–1019. DOI: [10.1016/j.crma.2010.08.003](https://doi.org/10.1016/j.crma.2010.08.003). URL: <http://dx.doi.org/10.1016/j.crma.2010.08.003>.
- [54] J. Friedman, T. Hastie and R. Tibshirani. 'Sparse inverse covariance estimation with the graphical lasso'. In: *Biostatistics* 9.3 (2008), pp. 432–441.
- [55] C. Giraud, S. Huet and N. Verzelen. 'Graph selection with GGMselect'. In: *Statistical applications in genetics and molecular biology* 11.3 (2012).
- [56] T. Hansen and U. Zwick. 'Lower Bounds for Howard's Algorithm for Finding Minimum Mean-Cost Cycles'. In: *ISAAC (1)*. 2010, pp. 415–426.
- [57] S. Herrmann and P. Vallois. 'From persistent random walk to the telegraph noise'. In: *Stoch. Dyn.* 10.2 (2010), pp. 161–196. DOI: [10.1142/S0219493710002905](https://doi.org/10.1142/S0219493710002905). URL: <http://dx.doi.org/10.1142/S0219493710002905>.
- [58] J. Hu, W.-C. Wu and S. Sastry. 'Modeling subtilin production in bacillus subtilis using stochastic hybrid systems'. In: *Hybrid Systems: Computation and Control*. Springer, 2004, pp. 417–431.
- [59] R. Keinj, T. Bastogne and P. Vallois. 'Multinomial model-based formulations of TCP and N TCP for radiotherapy treatment planning'. In: *Journal of Theoretical Biology* 279.1 (June 2011), pp. 55–62. DOI: [10.1016/j.jtbi.2011.03.025](https://doi.org/10.1016/j.jtbi.2011.03.025). URL: <http://hal.inria.fr/hal-00588935/en>.
- [60] R. Koenker. *Quantile regression*. 38. Cambridge university press, 2005.

- [61] Y. A. Kutoyants. *Statistical inference for ergodic diffusion processes*. Springer Series in Statistics. London: Springer-Verlag London Ltd., 2004, pp. xiv+481.
- [62] C. Lacaux. ‘Real Harmonizable Multifractional Lévy Motions’. In: *Ann. Inst. Poincaré*. 40.3 (2004), pp. 259–277.
- [63] B. Lalloué, J.-M. Monnez and E. Albuissou. ‘Convergence d’un score d’ensemble en ligne : étude empirique’. In: *52e Journées de Statistique*. Société Française de Statistique. Nice, France, July 2020. URL: <https://hal.archives-ouvertes.fr/hal-02894908>.
- [64] L. Lebart. ‘On the Benzecri’s method for computing eigenvectors by stochastic approximation (the case of binary data)’. In: *Compstat 1974 (Proc. Sympos. Computational Statist., Univ. Vienna, Vienna, 1974)*. Vienna: Physica Verlag, 1974, pp. 202–211.
- [65] B. Lesner and B. Scherrer. ‘Non-Stationary Approximate Modified Policy Iteration’. In: *ICML 2015*. Lille, France, July 2015. URL: <https://hal.inria.fr/hal-01186664>.
- [66] T. Lyons and Z. Qian. *System control and rough paths*. Oxford mathematical monographs. Clarendon Press, 2002. URL: <http://books.google.com/books?id=H9fRQNIgZYC>.
- [67] N. Meinshausen and P. Bühlmann. ‘High-dimensional graphs and variable selection with the lasso’. In: *The Annals of Statistics* (2006), pp. 1436–1462.
- [68] J.-M. Monnez. ‘Approximation stochastique en analyse factorielle multiple’. In: *Ann. I.S.U.P* 50.3 (2006), pp. 27–45.
- [69] J.-M. Monnez. ‘Convergence d’un processus d’approximation stochastique en analyse factorielle’. In: *Publ. Inst. Statist. Univ. Paris* 38.1 (1994), pp. 37–55.
- [70] J.-M. Monnez. ‘Stochastic approximation of the factors of a generalized canonical correlation analysis’. In: *Statist. Probab. Lett.* 78.14 (2008), pp. 2210–2216. DOI: [10.1016/j.spl.2008.01.088](https://doi.org/10.1016/j.spl.2008.01.088). URL: <http://dx.doi.org/10.1016/j.spl.2008.01.088>.
- [71] E. Nadaraya. ‘On non-parametric estimates of density functions and regression curves’. In: *Theory of Probability & Its Applications* 10.1 (1965), pp. 186–190.
- [72] I. Post and Y. Ye. *The simplex method is strongly polynomial for deterministic Markov decision processes*. Tech. rep. arXiv:1208.5083v2, 2012.
- [73] M. Puterman. *Markov Decision Processes*. Wiley, New York, 1994.
- [74] B. Roynette, P. Vallois and M. Yor. ‘Brownian penalizations related to excursion lengths, VII’. In: *Annales de l’IHP Probabilités et statistiques*. Vol. 45. 2. 2009, pp. 421–452.
- [75] F. Russo and P. Vallois. ‘Elements of stochastic calculus via regularization’. In: *Séminaire de Probabilités XL*. Vol. 1899. Lecture Notes in Math. Berlin: Springer, 2007, pp. 147–185. DOI: [10.1007/978-3-540-71189-6\\_7](https://doi.org/10.1007/978-3-540-71189-6_7). URL: [http://dx.doi.org/10.1007/978-3-540-71189-6\\_7](http://dx.doi.org/10.1007/978-3-540-71189-6_7).
- [76] F. Russo and P. Vallois. ‘Stochastic calculus with respect to continuous finite quadratic variation processes’. In: *Stochastics: An International Journal of Probability and Stochastic Processes* 70.1-2 (2000), pp. 1–40.
- [77] B. Scherrer. ‘Approximate Policy Iteration Schemes: A Comparison’. In: *ICML - 31st International Conference on Machine Learning - 2014*. Pékin, China, June 2014. URL: <https://hal.inria.fr/hal-00989982>.
- [78] B. Scherrer. ‘Improved and Generalized Upper Bounds on the Complexity of Policy Iteration’. In: *Mathematics of Operations Research* (Feb. 2016). Markov decision processes ; Dynamic Programming ; Analysis of Algorithms. DOI: [10.1287/moor.2015.0753](https://doi.org/10.1287/moor.2015.0753). URL: <https://hal.inria.fr/hal-00829532>.
- [79] B. Scherrer. ‘Performance Bounds for Lambda Policy Iteration and Application to the Game of Tetris’. In: *Journal of Machine Learning Research* 14 (Jan. 2013), pp. 1175–1221. URL: <https://hal.inria.fr/hal-00759102>.
- [80] B. Scherrer, M. Ghavamzadeh, V. Gabillon, B. Lesner and M. Geist. ‘Approximate Modified Policy Iteration and its Application to the Game of Tetris’. In: *Journal of Machine Learning Research* 16 (2015). A paraître, pp. 1629–1676. URL: <https://hal.inria.fr/hal-01091341>.

- [81] B. Scherrer and B. Lesner. ‘On the Use of Non-Stationary Policies for Stationary Infinite-Horizon Markov Decision Processes’. In: *NIPS 2012 - Neural Information Processing Systems*. South Lake Tahoe, United States, Dec. 2012. URL: <https://hal.inria.fr/hal-00758809>.
- [82] P. Vallois. ‘The range of a simple random walk on  $Z$ ’. In: *Advances in applied probability* (1996), pp. 1014–1033.
- [83] P. Vallois and C. S. Tapiero. ‘Memory-based persistence in a counting random walk process’. In: *Phys. A*. 386.1 (2007), pp. 303–307. DOI: [10.1016/j.physa.2007.08.027](https://doi.org/10.1016/j.physa.2007.08.027). URL: <http://dx.doi.org/10.1016/j.physa.2007.08.027>.
- [84] N. Villa-Vialaneix. *An introduction to network inference and mining*. <http://wikistat.fr/>. (consulté le 22/07/2015). 2015. URL: [http://www.nathalievilla.org/doc/pdf/wikistat-network\\_compiled.pdf](http://www.nathalievilla.org/doc/pdf/wikistat-network_compiled.pdf).
- [85] Y. Ye. ‘The Simplex and Policy-Iteration Methods Are Strongly Polynomial for the Markov Decision Problem with a Fixed Discount Rate’. In: *Math. Oper. Res.* 36.4 (2011), pp. 593–603.